



Luo, L., Clewer, DR., Canagarajah, CN., & Bull, DR. (2001). Genetic stereo matching using complex conjugate wavelet pyramids. In *IEEE International Conference on Image Processing (ICIP), Greece, October, 2001* (Vol. 2, pp. 153 - 156). Institute of Electrical and Electronics Engineers (IEEE).
<https://doi.org/10.1109/ICIP.2001.958447>

Peer reviewed version

Link to published version (if available):
[10.1109/ICIP.2001.958447](https://doi.org/10.1109/ICIP.2001.958447)

[Link to publication record in Explore Bristol Research](#)
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

GENETIC STEREO MATCHING USING COMPLEX CONJUGATE WAVELET PYRAMIDS

L. J. Luo*, D. R. Clewer, C. N. Canagarajah, and D. R. Bull

Centre for Communications Research, Univ. of Bristol, Bristol BS8 1UB, UK

*Oxford Semiconductors Ltd., 69 Milton Park, Abingdon, OX14 4RX, UK

ABSTRACT

In this paper, a new genetic algorithm-based optimisation technique for stereo matching using complex conjugate wavelet pyramids is proposed. Reliable disparity fields are estimated in the wavelet domain with low computational cost. The new cost function is composed of the differences in wavelet coefficient values, plus vertical discontinuity and ordering constraints. Within homogenous regions, smoothness constraints on the disparity field are also employed. A genetic algorithm is used, where previously estimated vectors at the former image hierarchy are used to predict the corresponding search space of chromosomes, and to correct each newly calculated set of disparity vectors. This significantly reduces computational complexity compared to other methods, whilst maintaining robust performance.

1. INTRODUCTION

The accurate computation of stereo depth is an important problem for many visual tasks. A very large number of algorithms have been proposed in the literature [1]. Approaches to the correspondence problem, namely disparity estimation, can be broadly classified into three categories: intensity-based or correlation-based matching, feature-based matching, and matching function optimization techniques. An alternative approach to intensity-based or correlation-based stereo matching, commonly known as the window-based method, is to only match those regions in the images that are interesting, for instance, regions that contain high variation of intensity values in the horizontal, vertical, and diagonal directions.

Feature-based matching is introduced naturally to overcome the inabilities of intensity-based or correlation-based matching by attempting matching only on information-rich points or more complicated primitives such as edges, regions, etc. The matching function optimization techniques find the depth fields that minimize some functions, usually call the energy or the objective function. These methods usually use some optimization technique, e.g., simulated annealing [2],

mean fields [3], graph cuts [4], neural networks [5], genetic algorithm [1], dynamic programming [6], etc. For global optimization methods, the major problem is the choice of energy function. The most natural energy functions for stereo are 2-D, and contain a data term and a smoothness term.

It is shown that wavelet multiresolution analysis provides an adequate transformation and representation of image signal information with desired properties such as good space-frequency locality and information preservation. Complex conjugate wavelets have been used for stereo disparity matching [7]. Stereo matching relies on exploiting the full information in the image pairs. As shown in [7], wavelet multiresolution analysis (wavelet pyramid) has been considered as a very good full-information representation of stereo image pairs for matching and it is uniform throughout scale space. Pan [7] uses complex wavelets to transform stereo image pairs into full-information pyramids. A global objective function is then established under the maximum *a posteriori* probability criterion and equivalently transformed into a minimum description length criterion.

In this paper, a new genetic algorithm-based optimisation technique for stereo matching using a wavelet pyramid (GSMWP) is proposed. Stereo matching is the essential process to recover three-dimensional structure of objects. The disparity is constructed as chromosomes with fitness values inversely proportional to their costs. The new cost function is composed of the wavelet-coefficient-difference between image pairs and smoothness constraints of disparity. The operations of the genetic algorithm are affected by the disparities of neighbouring pixels. Experimental results for various test images show that the proposed algorithm has very good performance.

This paper is organised as follows. The object function is firstly given in section 2. Section 3 describes the new genetic stereo matching algorithm using a wavelet pyramid. Section 4 gives experimental results from our investigations. Lastly a discussion and conclusions are presented in section 5.

2. THE OBJECT COST FUNCTION

We propose a new solution of the object cost function to stereo matching by combining a full-information representation of the complex conjugate wavelet pyramid, a vertical discontinuity constraint, a smoothness constraint, and an order constraint into a pixel-based method. More details of the wavelet pyramid decomposition can be found in [7]. Pixels in correspondence may be neither constrained by intensity values nor constrained by intensity gradient. Within a homogenous region, smoothness constraints are needed to make proper interpolations.

Suppose that $W_l(x, y)$ and $W_r(x, y)$ are the wavelet coefficients (luminance and colour) of point (x, y) in the wavelet pyramids of the left and right images. If the disparity at position (x, y) in the left image is denoted by $d_{x,y}$, from the intensity preservation principle, it follows that $W_l(x, y) = W_r(x + d_{x,y}, y)$. However, since such intensity measurements are not exactly fulfilled, the following cost function (including a smoothness constraint, a vertical-continuity constraint, and an ordering constraint) is employed for minimization.

$$e(i, j) = \underbrace{\sum_D |W_l(x, y) - W_r(x + d_{i,j}, y)|}_{\text{matching-error}} + \beta \underbrace{\sum |\nabla^2 d_{i,j}|}_{\text{smoothness-error}} \quad (1)$$

$$+ \gamma \underbrace{\sum_p |d_{i,j} - d_p|}_{\text{vertical-discontinuity-error}} + \lambda \underbrace{[f(d_{i-1,j} - d_{i,j} - 1) + f(d_{i,j} - d_{i+1,j} - 1)]}_{\text{ordering-constraint-error}}$$

where

$$\nabla^2 d_{i,j} = (d_{i+1,j} + d_{i-1,j} + d_{i,j+1} + d_{i,j-1} - 4d_{i,j}) \quad (2)$$

$$f(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (3)$$

$$N_p = \{(i, j - 2), (i, j - 1), (i, j), (i, j + 1), (i, j + 2)\} \quad (4)$$

A reference window D of size $p \times q$ is placed around point (i, j) , and then compared with a search window in a given horizontal interval. Constraints on the local shape of the disparity surface at point (x, y) also include the vertical-continuity constraint and the ordering constraint. When a step edge is caused by a false vertical discontinuity, the vertical-continuity constraint will help to smooth out the false vertical disparity edge.

The ordering-constraint check has been used in [8] and is violated if

$$d_{i-1,j} - d_{i,j} > 1 \quad (5)$$

The constraints used in equation (1) will add a penalty term to the cost function such that a global ordering is achieved. The selection of constant coefficients β, γ, λ will be discussed later.

3. THE GENETIC STEREO MATCHING PROCEDURE

Because of their robustness and effectiveness for efficiently solving traditionally difficult or multimodal problems, genetic algorithms have been successfully applied to various optimization problems since their theory was developed by Holland [9]. The GA consists of a string-representation of points called chromosomes in the search space, a fitness function to evaluate the search point, a set of operators for generating new chromosomes, and a stochastic assignment to control the genetic operators. In case of the disparity estimation, the GA is applied to refine the displacement field. A population P is maintained which consists of N search points along the epipolar line, where N is the population size. The population P evolves into another population P' in response to the application of certain genetic operations. Chromosomes with higher fitness values will have a higher probability of being kept in the population of the next generation, and hence propagating their offspring. On the other hand, weak chromosomes, whose fitness is small, will be replaced by new stronger chromosomes. Therefore, the quality of the chromosomes in the population will be improved. When the iteration process converges, the improved disparity vector is expected to be contained in the mature population.

In order to reduce noise sensitivity and simultaneously achieve higher efficiency, both the left image and right image are transformed into wavelet pyramids. In the first level in the image hierarchy, the search range is set to $(-1, 3)$ for the left image and $(-3, +1)$ for the right image, and the initial disparity vectors are all set to zero. At the h_{th} level of the wavelet pyramid, the initial disparity field is upsampled by a factor of two in both directions from that at the $(h-1)_{th}$ level of the wavelet pyramid. The above procedure repeats until the highest hierarchy level is reached. This incremental scheme offers significant reduction in computational complexity. Our proposed GSMWP algorithm therefore consists of the following steps: initialization, evaluation, selection, crossover and mutation. At each level in the image hierarchy, the GA algorithm progresses point by point from the top left to the bottom right in the hierarchical left and right images. For each point in an image, the chromosome with maximum fitness value is selected from the current population as the possible solution, namely the disparity vector. The whole process for a whole hierarchical image repeats until convergence (namely the matching error of the possible solution is less than a predefined threshold) or a predefined number of iterations have been performed.

4. EXPERIMENTAL RESULTS

The robustness of our proposed approaches has been tested on real stereoscopic pairs: AQUA and SANTA CLAUS recorded with different camera set-ups. The original image resolutions are 720x576 pixels and 384x288 pixels respectively. These sequences have been recorded using stereo cameras with baselines of 8.75cm (AQUA) and 10cm (Santa Claus). At the h_{th} level of image hierarchy of the GA method, the constraint parameters in equation (1) are selected empirically as follows

$$\beta = 5\eta; \gamma = 5\eta; \lambda = \eta$$

$$\text{where } \eta = \frac{p_h \cdot q_h}{p_{V-2} \cdot q_{V-2}}, h = 0, \dots, V-2 \quad (6)$$

Based on simulation results, the size $p \times q$ of the reference window D at the h_{th} level of image hierarchy of the GA method is chosen to be $p_h = q_h = V - 2 - h$. We have seen a sharp disparity map can be obtained using this method. Constraints on the disparity gradient function serve as a wavelet-coefficient-adaptive smoothness constraint.

Some results illustrating the performance of the HGA method [1] and the GSMWP method are given in Fig. 1. Images representing the horizontal component of dense L-R disparity fields for AQUA are shown. Here darker gray levels represent larger negative horizontal vectors, whereas brighter gray levels represent larger positive horizontal vectors. For our methods, the matching produces encouraging results. Except for the possibility of blurring of the occluded disparity edges, both of the proposed GSMWP and the HGA method produce very accurate disparity estimation for most flat or non-flat regions and it looks that the GSMWP method provides slightly better result than the HGA method.



(a) The HGA method; (b) The GSMWP method

Fig. 1 The left-right disparity field for AQUA

5. COMPARISON AND CONCLUSIONS

As mentioned in [1], performance comparisons between different disparity estimation algorithms are very difficult. High subjective quality of synthesized views when viewed individually is not sufficient to guarantee

high stereoscopic image quality. Artefacts present in the stereo image may not be obviously apparent when viewing each view monoscopically. Depth discontinuities and other stereo artefacts may be identified by the following methods:

(1) Comparing synthesized stereo image pairs with those produced by ground-truth disparity maps.

(2) Performing a "matching test", whereby straight vertical lines overlaid on the left image are deformed by the estimated disparity map and projected onto the right image. This gives an indication of incorrect disparity estimation and its resulting depth discontinuities.

(3) Displaying the set of synthesized images as an animation, where each image is displayed one after the other, from left to right. This gives the impression of a single camera moving past the objects in the scene, and can reveal depth discontinuities by their differing rate of change of perspective relative to their surroundings.

(4) Viewing the synthesized image pairs on a stereoscopic or multiview display. Again, this reveals depth discontinuities and other artefacts not obvious from analysing each view independently.

All the above methods can detect the presence of artefacts in the synthesized image pairs, but comparing the severity of artefacts produced by different disparity estimation algorithms is very difficult – quantifying these comparisons even more so. Perhaps the only effective way to compare the results from different methods is to hold subjective evaluation trials, presenting the images on a stereoscopic or multiview display to a set of non-expert viewers, and recording their mean opinion scores.

Szeliski [10] proposed two methodologies for comparison. The first relies upon ground truth and he made use of a stereo pair of multiview images from the University of Tsukuba with dense ground truth computed by hand. The second relies upon the notion of prediction error, which is the ability of a disparity map to predict an unseen third image, taken from a known camera position with respect to the input pair. In this paper, we use the first methodology to compare our proposed algorithm with some other algorithms.

Our proposed GSMWP method belongs to global methods based on energy minimization. We select five of them for comparison, they are the cooperative method [11], simulated annealing [2], mean fields [3], graph cuts [4], and genetic algorithm [1], along with the squared absolute difference (SAD) method and the normalised correlation method [10]. Fig. 2 shows the depth computed by those algorithms as well as the ground truth.

In Fig. 2, you can see that the graph cut method does an excellent job of filling in uniform intensity areas, and makes almost no errors in the low-textured areas in the image. Simulated annealing and mean-field estimation of global optimization methods, the SAD method of

intensity-based methods, and the normalised correlation method of the correlation methods seem to have comparable performance.

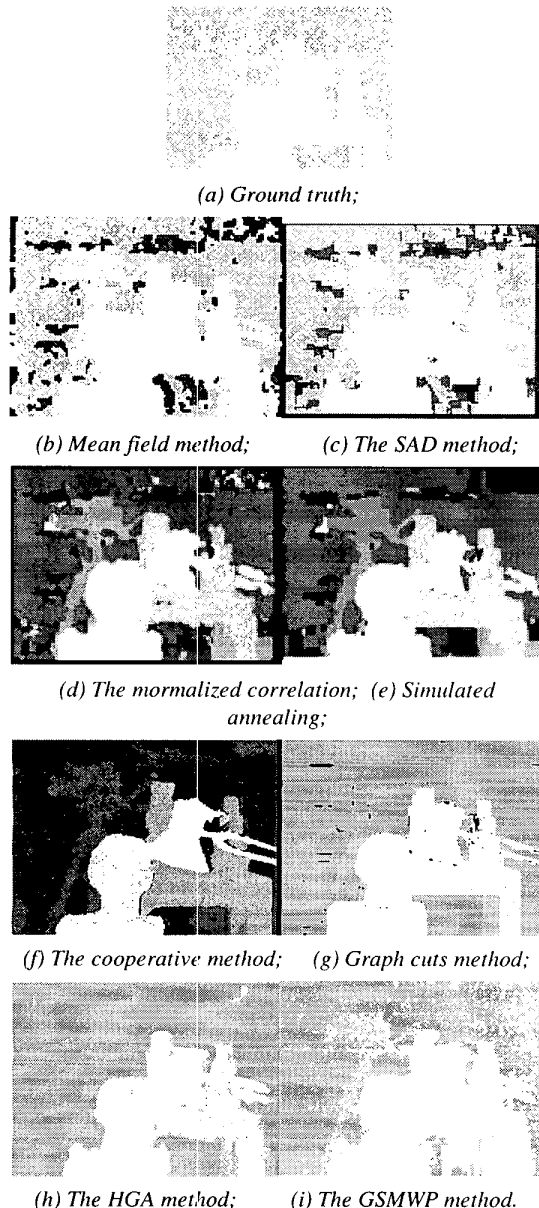


Fig. 2 Results for the imagery from Univ. of Tsukuba

Although the cooperative method can produce very good results, unfortunately, discretizing space volumetrically introduces a large number of degrees of

freedom and leads to sampling and aliasing artefacts. Our proposed GSMWP method can provide very good results for most of areas in the image. However, it does not look good in some flat regions. For this reason, future work will focus on how to improve the disparity vector in flat region. Thus we can see that all of the methods suggested so far have their limitations. Still, a tremendous amount of progress has been made in recent years in obtaining better disparity fields, intermediate view interpolation, and scene reconstruction.

6. ACKNOWLEDGEMENTS

This work was supported by the Virtual Centre of Excellence in Digital Broadcasting and Multimedia Technology Ltd. Acknowledgement also goes to the University of Tsukuba, Japan for use of their imagery in this paper.

REFERENCES

- [1] L. J. Luo, D. R. Clewer, D. R. Bull, C. N. Canagarajah, "A Hierarchical Genetic Disparity Estimation Algorithm for Multiview Synthesis", *Proc. 2000 IEEE Int. Conf on Image Processing*, Vancouver, Canada, Sept. 10-13, 2000.
- [2] S. Barnard, "Stochastic stereo matching over scale," *International journal of Computer Vision*, vol. 3, no.1, pp. 17-32, 1989.
- [3] D. Scharstein, and R. Szeliski, "Stereo matching with non-linear diffusion," *International journal of Computer Vision*, vol. 28, no.2, pp. 155-174, July 1998.
- [4] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *Seventh International Conference on Computer Vision*, Kerkyra, Greece, Sept. 1999.
- [5] R. D. Henkel, "A Simple and Fast Neural Network Approach to Stereovision," <http://axon.physik.uni-bremen.de/research/papers/>
- [6] Y. Ohta and T. Kanade, "Stereo by inter- and inter-scanline search using dynamic programming," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 7, no. 2, pp. 139-154, March 1995.
- [7] H. Pan, "Uniform full-information image matching using complex conjugate wavelet pyramids," *German Journal of Photogrammetry and Remote Sensing*, vol. 64, pp. 93-100, March 1996.
- [8] J. R. Ohm, and E Izquierdo, "An Object-Based System for Stereoscopic Viewpoint Synthesis," *IEEE Trans. on Circuits and Systems for Video Technology*, pp. 801-811, vol. 7, no. 5, October 1997.
- [9] H. Holland, *Adaptation in Natural and Artificial Systems*, Ann Arbor: Univ. Of Michigan, 1995.
- [10] R. Szeliski, and R. Zabih, "An experimental comparison of stereo algorithm," In *Vision Algorithms 99 Workshop*, Kerkyra, Greece, September 1999.
- [11] C. L. Zitnick and T. Kanade, "A volumetric iterative approach to stereo matching and occlusion detection," CMU-RI-TR-98-30.